



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
-----------------	-------------	----------------------	---------------------	------------------

10/608,590

06/27/2003

James M. Sweet

D/A2555Q

8455

25453 7590 04/12/2007

PATENT DOCUMENTATION CENTER

XEROX CORPORATION

100 CLINTON AVE., SOUTH, XEROX SQUARE, 20TH FLOOR

ROCHESTER, NY 14644

EXAMINER

HILLERY, NATHAN

ART UNIT

PAPER NUMBER

2176

SHORTENED STATUTORY PERIOD OF RESPONSE	MAIL DATE	DELIVERY MODE
--	-----------	---------------

3 MONTHS

04/12/2007

PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

<b>Office Action Summary</b>	<b>Application No.</b>	<b>Applicant(s)</b>	
	10/608,590	SWEET ET AL.	
	<b>Examiner</b>	<b>Art Unit</b>	
	Nathan Hillery	2176	

**-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --**

**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 05 February 2007.
- 2a) ☒ This action is **FINAL**.                      2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1-48 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-48 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All    b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- |   |   |
|---|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)                        | 4) <input type="checkbox"/> Interview Summary (PTO-413)                     |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)    | Paper No(s)/Mail Date. _____  |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152) |
| Paper No(s)/Mail Date _____   | 6) <input type="checkbox"/> Other: _____                                    |

### DETAILED ACTION

1. This action is responsive to communications: Amendment filed on 2/5/07.
2. Claims 1 – 48 are pending in the case. Claims 1, 26 and 37 are independent.

#### ***Claim Rejections - 35 USC § 101***

3. 35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

4. Claims 1 – 48 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter. Claims 1 – 48 have no practical application of a judicial exception as claimed because there is no physical transformation and no production of a concrete, useful and tangible result.

- a. The claimed invention remains in the abstract and nothing is made available to the user; thus it does not produce a tangible result.
- b. The claims appear to be in the preliminary stages and fall short of the disclosed practical utility. In other words, the claims fail to fulfill and/or reflect the specific, substantial, and credible utility sought by the disclosed invention, and thus do not produce a useful result.

5. Consequently, the claims are nonstatutory. The claims simply recite methodologies for assembling and grouping data without producing a concrete, useful, and tangible result.

6. Further, to expedite a complete examination of the instant application the claims rejected under 35 U.S.C. 101 (nonstatutory) above are further rejected as set forth below in anticipation of applicant amending these claims to make them statutory.

***Claim Rejections - 35 USC § 112***

7. The following is a quotation of the first paragraph of 35 U.S.C. 112:

The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full, clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same and shall set forth the best mode contemplated by the inventor of carrying out his invention.

8. Claims 1 – 42 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention.

9. The original specification as filed provides no support for **a document representation stored in memory** (penultimate lines of claims 1, 26 and 37).

10. Claims 2 – 25, 27 – 36, and 38 – 42, the claims are rejected for fully incorporating all of the deficiencies of the base claim(s) from which they depend.

***Claim Rejections - 35 USC § 103***

11. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

12. Claims 1 – 6, 10, 12 – 18, 26 – 30, 35 – 41, and 46 – 48 are rejected under 35 U.S.C. 103(a) as being unpatentable over Bharat et al. (US 6112203 A) and in further view of Earl (US 5924104 A).

13. **Regarding independent claim 37**, Bharat et al. teach that we locate pages that point to at least one of the pages in the start set 201. We call this set of pages the back set 202 (Column 4, line 61 – Column 5, line 20), which meets the limitation of **performing a page-level link analysis that identifies those hyperlinks on a page linking to a candidate document page.**

Bharat et al. teach that if a link points to a page that is represented by a node in the graph, and both pages are on different servers, then a corresponding edge 213 is added to the graph 211. Nodes representing pages on the same server are not linked. This prevents a single Web site with many self-referencing pages to unduly influence the outcome (Column 4, line 61 – Column 5, line 20), which meets the limitation of **identifying possible progression links; identifying possible table of content links; and examining the possible progression links and the possible table of content links for common characteristics**, since the specification states that “the possible progression links 230 and possible table of content links 240 are passed to module 250 for a final examination to weed out links which have properties that are not characteristic of typical intra-document links, e.g. they point to a different web server” (p 7, lines 26 – 30). It should be noted that pages on the same server are nodes and are thus still apart of the resulting graph.

Furthermore, it should be noted that the self-referencing pages of Bharat et al. are equivalent to intra-document links and that those intra-document links can be “possible” progression and/or table of contents links, since the Office has interpreted the word “possible” as “could be” and within the broadest, reasonable interpretation in light

of the specification, which states that a link analysis phase consists of the identification for a given hypertext page of the most likely desirable intra-document links. Those intra-document links fall into two categories: progression links and table of contents links (p 5, second paragraph). Thus, any intra-document link – a link that points to a different web server – could be a possible progression or table of contents link.

Bharat et al. teach that a larger n-graph 211 can be constructed by repeating this process for the back and forward sets 202-203 to add more indirectly linked pages (Column 4, line 61 – Column 5, line 20), which meets the limitation of **performing a recursive application of the page-level link analysis to the linked candidate document page and any further nested candidate document pages thereby identified, until a collective set of identified candidate document pages is assembled.**

Bharat et al. do not explicitly teach **performing a document-level analysis that examines the collective set of identified candidate document pages for grouping into one or more documents; examining the collective set of identified candidate document pages to weed out links which have properties that are not characteristic of typical intra-document links, to provide a resultant set of identified candidate document pages; and grouping the content found in the resultant set of candidate document pages into a document representation for subsequent viewing or printing of the given hyperdocument.**

Earl teaches that the link display manager 300 includes a document parser 304 for parsing each document and identifying links 202 and 204 (Column 2, line 59 – Column 3, line 9), which meets the limitation of **performing a document-level analysis that examines the collective set of identified candidate document pages for grouping into one or more documents.**

Earl teaches that the link display manager 300 includes a display system for defining predetermined screen element properties providing visual cues for distinguishing the identified links 202 and 204. When a user provides an input link selection to select a new document, the document parser 304 parses the selected new document to identify intradocument links 202 and interdocument links 204 (Column 2, line 59 – Column 3, line 9), which meets the limitation of **examining the collective set of identified candidate document pages to weed out links which have properties that are not characteristic of intra-document links, to provide a resultant set of identified candidate document pages.**

Earl teaches that the display system 306 processes the identified intradocument links 202 and interdocument links 204 for displaying distinctively the intradocument links 202 and interdocument links 204 with predetermined visual cues to differentiate the links 202, 204 (Column 2, line 59 – Column 3, line 9), which meets the limitation of **grouping the content found in the resultant set of candidate document pages into a document representation stored in memory for subsequent viewing or printing by a user of the given hyperdocument.**

It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the invention of Bharat et al. with that of Earl because such a combination would provide the users of Bharat et al. with *an improved method and apparatus for displaying links on a user display interface in a computer system* (Column 1, lines 39 – 41).

14. **Regarding dependent claims 38 – 41**, Bharat et al. teach that the nodes in the start set are first scored according to their connectivity, and the number of terms of the query that appear as unique sub-strings in the URL of the represented documents. The score is a weighted sum of the number of directed edges to and from a node and the number of unique sub-strings of the URL that match a query term (Column 3, lines 10 – 15), meets the limitation of **the page-level link analysis includes examination of contextual clues, the contextual clue is a particular class of content item associated with the hyperlink, the class of content item is a class of text, the class of text is a directional word or phrase.**

15. **Regarding dependent claim 46**, Bharat et al. teach that we assign a similarity weight to each node 213 of the sub-graph 255. Various document similarity measuring techniques have been developed in Information Retrieval to determine the goodness of fit between a "target" document and a collection of documents. These techniques typically measure a similarity score based on word frequencies in the collection and a target document (Column 6, lines 51 – 57), meets the limitation of **the contextual clue**



**is the similarity of the hyperlink destination to that of other hyperlinks within the hyperdocument.**

16. **Regarding dependent claim 47**, Bharat et al. teach that we use a modified Kleinberg algorithm on the nodes of the pruned n-graph 265 to determine useful hub and authority pages. For each node of the pruned n-graph 265, we measure two scores: a hub score (HS), which estimates how good a hub the page is, and an authority score (AS), which estimates how good an authority the page is. The intuition behind our method is this: a good hub is one that points to many documents. A good authority is one that is pointed to by many documents. Transitively, an even better hub is one that points to many good authorities, and an even better authority is one that is pointed to by many good hubs (Column 7, lines 41 – 50), meets the limitation of **the document-level analysis includes the identification of pages forming a chain of progression links.**

17. **Regarding dependent claims 18 and 48**, Bharat et al. teach that after we have constructed the nodes 212, we add the directed edges 213. If a link points to a page that is represented by a node in the graph, and both pages are on different servers, then a corresponding edge 213 is added to the graph 211. Nodes representing pages on the same server are not linked. This prevents a single Web site with many self-referencing pages to unduly influence the outcome. This completes the n-graph 211 (Column 5, lines 13 – 20), meets the limitation of **the similarity includes the location**

**at which the page is stored, and the document-level analysis includes the identification of pages linked to by the same tables of contents.**

18. **Regarding claims 1 – 6, 10 and 12 – 14**, the claims incorporate substantially similar subject matter as claims 37 – 41 and 46 – 48 and are rejected along the same rationale.

19. **Regarding claims 26 – 30, 35 and 36**, the claims incorporate substantially similar subject matter as claims 37 – 41 and 46 – 48 and are rejected along the same rationale.

20. **Regarding dependent claim 15 – 17**, Bharat et al. teach that we use do iterative connectivity analysis 310, content analysis 320, and pruning 330. This method consists of a sequence of rounds. In each round, our modified connectivity analysis is run for 10 iterations to get a listing of the (current) best hubs and authorities 315. In step 320, the pages are examined for content similarity in decreasing order of rank, alternating between the hub and the authority list. Less relevant pages are pruned (Column 8, lines 25 – 33), meets the limitation of **the document-level analysis includes identifying the pages listed in a table of contents, the document-level analysis includes identifying as part of the document the page containing the table of contents, the document-level analysis includes the similarity of candidate pages.**

Art Unit: 2176

21. Claims 21 and 22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Bharat et al. (US 6112203 A) and Earl (US 5924104 A) as applied to claims 1 – 6, 10, 12 – 18, 26 – 30, 35 – 41, and 46 – 48 above, and further in view of Huang et al. (US 6601075 B1).

22. **Regarding dependent claims 21 and 22**, neither Bharat et al. nor Earl teach **the similarity includes similar style specifications, and the similarity includes similar page layout.**

Huang et al. teach that the HITS and CLEVER algorithms make use of hyperlinked structures to rank documents that share the same schema. Exemplary documents with hyperlinked structures are HTML documents. XML has given rise to a new hyperlink environment that includes documents with different schemas. Hence, this new environment helps rank documents based on the quality of their associated schema, determine the quality of the schemas themselves, and ranks documents based on their structural properties (e.g. validity, well-formedness, etc.) (Column 3, lines 37 – 53), meets the limitation of **the similarity includes similar style specifications, and the similarity includes similar page layout.**

It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the combined invention of Bharat et al. and Earl with that of Huang et al. because such a combination would allow the users of Bharat et al. and Earl the benefit of *an algorithm which is applied to an initial set of documents, similar to the HITS and CLEVER algorithms* (Column 3, line 66 – Column 4, line 1).

Art Unit: 2176

23. Claims 23 and 25 are rejected under 35 U.S.C. 103(a) as being unpatentable over Bharat et al. (US 6112203 A) and Earl (US 5924104 A) as applied to claims 1 – 6, 10, 12 – 18, 26 – 30, 35 – 41, and 46 – 48 above, and further in view of Law et al. (US 6754873 B1).

24. **Regarding dependent claims 23 and 25, neither Bharat et al. nor Earl teach the similarity includes similar logical structure of the page content, the document-level analysis includes analysis of the topological structure of the linked pages.**

Law et al. teach that the link structure of the hyperlinked documents is analyzed in order to find hyperlinked documents that are related to and at the same level of generality of a hyperlinked document (Column 2, lines 8 – 11), meets the limitation of **the similarity includes similar logical structure of the page content, the document-level analysis includes analysis of the topological structure of the linked pages.**

It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the combined invention of Bharat et al. and Earl with that of Law et al. because such a combination would allow the users of Bharat et al. and Earl the benefit of *innovative techniques for finding related hyperlinked documents using link-based analysis* (Column 2, lines 6 – 8).

25. Claims 7 – 9, 11, 19, 20, 24, 31 – 34, and 42 – 45 are rejected under 35 U.S.C. 103(a) as being unpatentable over Bharat et al. (US 6112203 A) and Earl (US

Art Unit: 2176

5924104 A) as applied to claims 1 – 6, 10, 12 – 18, 26 – 30, 35 – 41, and 46 – 48 above, and further in view of Prince (US 6877002 B2).

26. **Regarding dependent claims 7 – 9, 11, 19, 20, 24, 31 – 34, and 42 – 45,** neither Bharat et al. nor Earl explicitly teach **meta-data** or **image**.

However, Prince teaches that the parsed results (from step 42 in FIG. 4) relating to the media are passed to extraction agent 68 via an extraction queue 67. The extraction queue 67 comprises URLs to be analyzed with respect to associated media metadata. The extraction queue 67 may comprise metadata queue entries such as media URLs, Web page URLs, Web page titles, Web page keywords, Web page descriptions, media title, media author, and media genre. Each queue entry added to the extraction queue is assigned a processing time and a priority (Column 7, lines 23 – 37), meets the limitation of **the similarity includes the similarity of meta-data associated with the page, the meta-data includes the author identification, the similarity includes the presence of at least one similar content item on each page, the class of content item is a class of image, the class of image is an image containing a directional symbol, a textual clue is obtained for the class of image, the contextual clue is the presence of at least one other hyperlink nearby with the candidate document page.**

It would have been obvious to one of ordinary skill in the art at the time of the invention to combine the combined invention of Bharat et al. and Earl with that of Prince because such a combination would allow the users of Bharat et al. and Earl the benefit

Art Unit: 2176

*of A method for querying metadata associated with media on a computer network includes separating the metadata into keywords (Column 2, lines 37 – 39).*

### ***Response to Arguments***

27. Applicant's arguments filed 2/5/07 have been fully considered but they are not persuasive.

28. Applicant argues that claims 1 – 48 are statutory under 35 USC 101 because a document stored in memory is not a judicial exception and a transformation occurs (p 10, last full paragraph – p 11, second full paragraph).

The Office disagrees.

First, a document stored in memory, whether inside someone's brain or on a computer, constitutes a judicial exception; specifically, it is considered an abstract idea. Secondly, the transformation to which applicant eludes, web page data into a document representation stored in memory, simply constitutes a data transformation not a physical transformation.

Further, applicant subtly requests suggestions on how to overcome the 101 rejection. To this end, it is suggested that applicant amend each independent claim to make them statutory by producing a tangible and useful result. For example, positively reciting that the document is printed, stored or displayed NOT insinuating that it might happen some time in the future and NOT simply reciting an intended use, future or otherwise, for the document. It should be noted that the suggestions mentioned, even if made, may not meet the standards under 35 USC 112, first paragraph.

29. Applicant argues that Earl does not teach **examining the collective set of identified candidate document pages to weed out links which have properties that are not characteristic of intra-document links, to provide a resultant set of identified candidate document pages** because what Earl defines as intra-document is what Applicant would call intra-page and thus what Earl calls inter-document is really inter-page (p 13, last paragraph).

The Office disagrees.

First, nowhere in the original Specification does Applicant define or even discuss "intra-page" or "inter-page". Second, Applicant proffers a "definition" of intra-document taken from the Specification (p 14, first paragraph). "The first step for an automated system for the identification of multi-page documents is to identify links within a given web page that may link to other pages within the same document. Such links are referred to as intra-document links". This "definition" is clearly NOT limiting in any way.

Second, the claim recites **links, which have properties that are not characteristic of intra-document links**. Consequently, the Office is forced to rely upon the knowledge of one of ordinary skill in the art in order to interpret the broad limitation recited in light of the lack of a definitive definition in the Specification. Thus, the Office maintains that Earl clearly and explicitly teaches intra-document and inter-document links that meet the claimed intra-document link in spite of Applicant's attempts to split hairs without sufficient evidence to support it.

Art Unit: 2176

30. Applicant argues that Earl further does not teach **examining the collective set of identified candidate document pages to weed out links which have properties that are not characteristic of intra-document links, to provide a resultant set of identified candidate document pages** because Earl discriminates between two types of links but keeps all those links while applicant discards or weeds out those links (p 14, third full paragraph).

The Office disagrees.

First, Applicant is correct that Encarta defines "weed out" as "to separate out something undesirable. However, the Office maintains that Earl does "weed out" the links within the broadest, reasonable interpretation in light of the specification. The term "weed out" is not defined in the specification. Although the applicant attempts to explain what the term "weed out" should mean as it pertains to gardening, the Office is forced to rely on the knowledge of one of ordinary skill in the art NOT of gardening but of computer technology.

Thus, Earl, by applicant's own admission, teaches discriminating visually between intra-document and inter-document links, which meet the definition of separating out, or weeding out, the links visually on screen. The requirement to have to discard the links is too limiting in view of what is actually claimed.

### ***Conclusion***

31. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP



§ 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

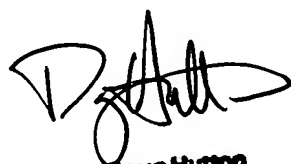
Any inquiry concerning this communication or earlier communications from the examiner should be directed to Nathan Hillery whose telephone number is (571) 272-4091. The examiner can normally be reached on M - F, 10:30 a.m. - 7:00 p.m.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Heather R. Herndon can be reached on (571) 272-4136. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Art Unit: 2176

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

NH



**Doug Hutton**  
Primary Examiner  
Technology Center 2100